Renewable and Sustainable Energy Reviews 63 (2016) 423–434

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Identifying the most significant input parameters for predicting global solar radiation using an ANFIS selection procedure

Kasra Mohammadi [a], Shahaboddin Shamshirband [b,*], Amirrudin Kamsin [b], P.C. Lai [c], Zulkefli Mansor [d]

[a] Department of Mechanical and Industrial Engineering, University of Massachusetts, Amherst, MA 01003, USA
[b] Department of Computer System and Technology, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia
[c] Institute of Research Management & Monitoring, Deputy Vice Chancellor (Research & Innovation), University of Malaya, Kuala Lumpur, Malaysia
[d] Research Center for Software Technology and Management, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 12 November 2015
Received in revised form
3 May 2016
Accepted 17 May 2016

Keywords:
Daily global solar radiation
ANFIS
Variable selection
Prediction
Solar energy systems

ABSTRACT

There are several variables that influence the global solar radiation (GSR) prediction; thus, determining the most significant parameters is an important task to achieve accurate predictions. In this paper, adaptive neuro-fuzzy inference system (ANFIS) is employed to identify the most relevant parameters for prediction of daily GSR. Three cities of Isfahan, Kerman and Tabass distributed in central and south central parts of Iran are considered as case studies. The ANFIS process for variable selection includes evaluating several combinations of input parameters for three cases with 1, 2 and 3 inputs to recognize the most relevant sets. To achieve this, nine parameters of sunshine duration (n), maximum possible sunshine duration (N), minimum, maximum and average air temperatures (T_{min}, T_{max} and T_{avg}), relative humidity (R_h), water vapor pressure (V_P), sea level pressure (P) and extraterrestrial radiation (H_o) are considered. The results reveal that an optimum sets of inputs are not identical for all cities due to difference in climate conditions and solar radiation characteristics. According to the results, considering the most relevant combinations of 2 input parameters is the more appropriate option for all cities to achieve more accuracy and less complexity in predictions. The survey results emphasize the importance of appropriate selection of input parameters to predict daily GSR. Such suitable, simple and accurate prediction is profitable to properly design and evaluate the performance of solar energy systems, which subsequently leads to technical and economic benefits.

Contents

* Corresponding author. Tel: +60 146266763.
E-mail addresses: kmohammadi@umass.edu (K. Mohammadi), shamshirband@um.edu.my (S. Shamshirband).

http://dx.doi.org/10.1016/j.rser.2016.05.065
1364-0321/

1. Introduction

Due to the destructive influences fossil fuels on the environment and rapid growth in the global energy demands, renewable energy sources are receiving further attention from researchers, governments and industries. Solar energy is a promising renewable energy source due to its unique characteristic, which is widely accessible around the world. Solar energy is utilized either directly or indirectly in a substantial number of applications such as electricity generation, air and water heating/cooling and etc. In fact, solar energy based technologies would play a substantial role in providing large portions of the current and future of global energy demands for both private and public sectors. Currently, solar energy harnessing is getting cheaper and more efficient due to technological changes, market competition and also enhancement in the capture, conversion and distribution processes. Solar energy harnessing would be highly profitable to mitigate the global warming and offer sustainability. The knowledge of global solar radiation is a fundamental requirement in different technological and scientific applications of solar energy such as appraising the performance of established solar technologies, as well as estimating the feasibility of installations in the future [1–4].

Ideally, the solar radiation information of every specific location should be acquired from measurement stations using the high precision instruments. Nonetheless, the reliable measured global solar radiation data are relatively scarce in many locations around the world because of paucity of instruments and fiscal issues [5,6]. As a consequence, several empirical models have been suggested for estimating global solar radiation using different meteorological and geographical elements [7–20]. Recently, soft-computing approaches have also been employed as particularly effective techniques to model global solar radiation in many parts of the globe.

Soft-computing techniques can efficiently recognize the relationships between dependent and independent variables even for non-linear natural processes. The adaptive neuro-fuzzy inference system (ANFIS) is a soft-computing technique which its employment brings many advantages such as adaptability for optimization and adaptive methods as well as computational efficiency. The main idea behind ANFIS is providing a method for the fuzzy modeling procedure to learn the information about data and organize the fuzzy inference system with given input/output data pairs.

Basically, there are several variables that have been typically used as required inputs to predict global solar radiation. Because considering many input parameters can increase the complexity and decrease the generalization of the models, determining the most significant variables is an important task to achieve accurate predictions. In this research work, an ANFIS selection procedure is applied to determine the most influential parameters affecting the horizontal global solar radiation. The main aim is to introduce the most significant parameters for prediction of daily horizontal global solar radiation in three cities of Isfahan, Kerman and Tabass distributed in the central and south central parts of Iran. These regions of Iran has been considered as case study owing to the remarkable solar energy potential. Also, measured global solar radiation data are rarely accessible. Nevertheless, the applied technique is totally applicable for other locations around the world. The process, named variable selection, includes a number of ways to determine a subset of the total recorded parameters that show favorable capability of prediction. The ANFIS network is utilized to examine how nine parameters (n, N, T_{min}, T_{max}, T_{avg}, R_h, V_P, P, and H_o) influence the daily global solar radiation.

The remainder of the article is structured as follows: In Section 2, a review of the related literature about the selection and importance of more relevant variables on solar radiation prediction is offered. Section 3 describes the considered case studies and data sets. In Section 4, the ANFIS methodology and variable selection process are explained. Afterwards the results and discussion are described in Section 5. Finally, the conclusions are offered in Section 6.

2. Literature survey

Due to existing correlations between solar radiation and other variables, many researchers have applied different combinations of input variables and analyzed their importance on accurate predictions of solar radiation, and then determined the more important ones. In this section, some of the important and related studies are reviewed in brief.

Tymvios et al. [21] compared the Angström-Prescott and artificial neural network (ANN) models for estimation of daily global solar radiation in Athalassa, Cyprus. They used the parameters of sunshine hour, maximum possible sunshine hour, number of month and maximum air temperature and developed 5 different ANN models; 2 models with 2 inputs and 3 models with 3 inputs. According to their results, considering sunshine hour, maximum possible sunshine hour and maximum air temperature as inputs provided higher precision. Bosch et al. [22] applied the automatic relevance determination (ARD) technique to select the relevant inputs for ANN model to estimate daily global radiation over complex terrains in south east of Spain. They assessed the significance of latitude, longitude, altitude, slope and azimuth angles, extraterrestrial solar radiation, clearness index and day of the year. They introduced three elements of altitude, day of the year and clearness index as the more appropriate inputs for ANN. Benghanem et al. [23] established six ANN-based models for daily global solar radiation estimation in Al-Madinah in Saudi Arabia. They utilized different combinations of input variables such as sunshine hour, ambient temperature, relative humidity and day of the year. Their results revealed that a higher accurate model was dependent on sunshine duration and air temperature. Behrang et al. [24] tested the performance of different ANN techniques to predict daily global solar radiation in Dezful, Iran. They considered different sets of meteorological parameters as inputs and showed that more precision was attained by considering five variables of day of the year, average air temperature, relative humidity, sunshine hours and wind speed as inputs. Koca et al. [25] used ANN technique to estimate global solar radiation for Mediterranean region of Anatolia in Turkey. They considered 6 different combinations of parameters with 4, 5 and 6 inputs elements and found that higher accuracy can be obtained by considering latitude, longitude, altitude, number of month, cloudiness and sunshine hour. Yacef et al. [26] assessed the performance of Bayesian Neural Network (BNN) in comparison with classical Neural Network (NN) and empirical models to estimate daily global solar irradiation in Madinah, Saudi Arabia. They used four different input elements of air temperature, relative humidity, sunshine duration and

Nomenclature

ANFIS	adaptive neuro-fuzzy inferences system
A, *B*, *C* and *D*	fuzzy set
a, *b* and *c*	membership function parameters
G_{sc}	solar constant (equal to 1367 W/m^2)
H	daily global radiation on horizontal surface (MJ/m^2)
H_o	daily extraterrestrial radiation on horizontal surface (MJ/m^2)
K_T	daily clearness index
MABE	mean absolute bias error (MJ/m^2)
n	daily sunshine hours (hr)
n_{day}	number of days
N	daily maximum possible sunshine hours (hr)
$O_{l,\ i}$	output of the *i*th node in layer *l*
pi, *qi* and *ri*	variable set
P	daily sea level pressure (mb)
R	correlation coefficient
RMSE	root mean square error (MJ/m^2)
R_h	daily relative humidity (%)
T_{avg}	daily average ambient temperature (°C)
T_{max}	daily maximum ambient temperature (°C)
T_{min}	daily minimum ambient temperature (°C)
VP	daily water vapor pressure (mb)

Greek letters

δ	solar declination angle (deg.)
ϕ	latitude of the location (deg.)
ω_s	sunset hour angle (deg.)
μ	membership function

extraterrestrial irradiation and employed Automatic relevance determination (ARD) method to determine the optimal input element of the NN. Their results showed that sunshine duration is the most relevant parameter followed by air temperature. Ozgoren et al. [27] developed an ANN model on the basis of multi-non-linear regression (MNLR) method to estimate the monthly global solar radiation over Turkey. They established 10 models using various variables by considering 1–10 input elements. They found that by considering 10 input parameters of sunshine hours, number of month, cloudiness, soil temperature, maximum, minimum and average atmospheric temperatures, wind speed, altitude and latitude gives higher precision. Mostafavi et al. [28] developed a hybrid approach by combining genetic programming (GP) with simulated annealing (SA) for global solar radiation estimation in two Iranian cities. They performed a sensitivity analysis to assess the influence of maximum and minimum air temperatures, relative humidity, precipitation and sunshine hours on global solar radiation estimation. Maximum and minimum air temperatures were found to be more relevant. Bhardwaj et al. [29] proposed a combined method using hidden Markov and generalized fuzzy for solar irradiation estimation in India. They assessed the influence of 15 different sets of meteorological parameters on solar radiation estimation. Their results specified that sunshine hour is the most significant element followed by air temperature, relative humidity, atmospheric pressure and wind speed, respectively. Ramedani et al. [30] developed 5 ANN-based models using 5 different sets of input data to predict daily global solar radiation in Tehran, Iran. They found that combination of maximum and minimum air temperatures, sunshine hour, maximum possible sunshine hour, extraterrestrial radiation and number of day offers more accuracy. Will et al. [31] utilized the Niching Genetic Algorithms (NGA) to identify the most important input elements for estimation of daily global solar radiation in El Colmenar (Tucumán, Argentina). They performed the analysis using measured data of 14 stations distributed in North part of Argentina. They considered different inputs elements consisting the maximum, minimum and average air temperatures, day of the year, relative humidity, atmospheric pressure, cloudiness and sunshine hour. Yadav et al. [32] employed Waikato Environment for Knowledge Analysis (WEKA) software for determining the most significant parameters in 26 locations of India to predict global solar radiation based on ANN. They used latitude, longitude, altitude, sunshine hours as well as maximum, minimum and average air temperatures as inputs. The latitude and longitude were determined as the least influential parameters while maximum, minimum and average air temperatures as well as altitude and sunshine hour were introduced as the most important inputs. Mohammadi et al. [33] developed a combined SVM-WT model using the support vector machine (SVM) and wavelet transform (WT) algorithm to predict both daily and monthly mean global solar radiation in an Iranian city. They used 3 different combinations of parameters and found that for both daily and monthly scales the combination of relative sunshine duration, difference between air temperatures, relative humidity, average temperature and extraterrestrial solar radiation as inputs, lead to attaining more accuracy. Also, the extraterrestrial solar radiation was recognized as a very important parameter. Yadav and Chandel [34] employed J48 algorithm using WEKA software to identify the most relevant input elements for ANN-based prediction of global solar radiation in western Himalayan Indian state of Himachal Pradesh. They introduced the air temperatures, altitude and sunshine duration as the most significant elements and recognized the latitude, longitude, clearness index and extra-terrestrial radiation as the least important parameters. They developed 5 ANN models to test the efficiency of J48 algorithm. Lopez et al. [35] applied the automatic relevance determination method (ARD) to determine more influential variables for estimation of hourly direct solar irradiance. By considering several atmospheric and radiometric variables, they found that clearness index and relative air mass are more important input variables to estimate direct solar irradiance using neural networks. Mohammadi et al. [36] applied ANFIS methodology to determine the most important variables for prediction of daily horizontal diffuse solar radiation in city of Kerman, Iran. They analyzed the influence of ten important variables and evaluated various combinations of input parameters. According to their results, for the cases with one input, sunshine duration (n) was determined as the most important variable. In addition, for the cases with 2 and 3 inputs, combination of horizontal global solar radiation (H) and extra-terrestrial solar radiation (H_o) as well as combination of H, H_o and n were recognized as the best sets for prediction of H_d, respectively.

In Table 1, the summary of conducted studies on the used and more relevant input parameters for prediction of different components of solar radiation is presented.

Reviewing the related literature reveals that owing to the dependency of solar radiation to different factors such as climate conditions and geographical location, it is not feasible to introduce an optimal or a set of optimum parameters with the highest impact on solar radiation prediction for all regions. Therefore, proper selections of more significant input parameters for prediction of solar radiation to provide more precision and less complexity would be of indispensable significance for every

Table 1
Summary of reviewed studies on the used and more relevant input parameters for solar radiation prediction.

Reference	Model	Location	Solar radiation component	Considered input parameters	More relevant input parameters
Tymvios et al. [21]	ANN	Athalassa, Cyprus	Global solar radiation	Five different input combinations using sunshine duration, maximum possible sunshine duration, number of month and maximum air temperature	Combination of sunshine duration, maximum possible sunshine duration and maximum air temperature
Bosch et al. [22]	ARD	12 stations in south east of spain	Global solar radiation	Seven different input combinations using latitude, longitude, altitude, slope and azimuth angles, extraterrestrial solar radiation, clearness index and day of the year	Combination of altitude, day of the year and clearness index
Benghanem et al. [23]	ANN	Al-Madinah, Saudi Arabia	Global solar radiation	Six different input combinations using sunshine duration, ambient temperature, relative humidity and day of the year	Combination of sunshine duration and air temperature
Behrang et al. [24]	ANN	Dezful, Iran	Global solar radiation	Six different input combinations using day of the year, daily mean air temperature, relative humidity, sunshine duration, evaporation and wind speed	Combination of day of the year, average air temperature, relative humidity, sunshine duration and wind speed
Koca et al. [25]	ANN	Seven cities located in Mediterranean region of Anatolia in Turkey	Global solar radiation	Six different input combinations using latitude, longitude, altitude, number of month, cloudiness, average temperature, relative humidity, wind velocity and sunshine duration	Combination of latitude, longitude, altitude, number of month, cloudiness and sunshine duration
Yacef et al. [26]	BNN	Madinah, Saudi Arabia	Global solar radiation	Average air temperature, relative humidity, sunshine duration and extraterrestrial solar radiation	Sunshine duration was identified as the most relevant parameter followed by air temperature
Ozgoren et al. [27]	ANN-MNLR	31 stations in Turkey	Global solar radiation	Ten different input combinations using latitude, longitude, altitude, number of month, monthly minimum, maximum and average air temperatures, soil temperature, relative humidity, wind speed, rainfall, atmospheric pressure, vapor pressure, cloudiness and sunshine duration	Combination of sunshine duration, number of month, cloudiness, soil temperature, maximum, minimum and average air temperatures, wind speed, altitude and latitude
Mostafavi et al. [28]	GP-SA	Two cities of Tehran and Kerman in Iran	Global solar radiation	maximum and minimum air temperatures, relative humidity, precipitation and sunshine duration	maximum and minimum air temperatures
Bhardwaj et al. [29]	hidden Markov-generalized fuzzy	Gurgaon, India	Global solar radiation	Sixteen different input combinations using sunshine duration, air temperature, relative humidity, wind speed and atmospheric pressure	Sunshine duration was determined as the most significant element followed by air temperature, relative humidity, atmospheric pressure and wind speed
Ramedani et al. [30]	ANN	Tehran, Iran	Global solar radiation	Five different input combinations using maximum, minimum and average air temperatures, sunshine duration, maximum possible sunshine duration, extraterrestrial solar radiation, precipitation and number of day	combination of maximum and minimum air temperatures, sunshine duration, maximum possible sunshine duration, extraterrestrial solar radiation and number of day
Will et al. [31]	NGA	14 stations in El Colmenar (Tucumán, Argentina	Global solar radiation	Several combinations using maximum, minimum and average air temperatures, day of the year, relative humidity, atmospheric pressure, cloudiness and sunshine duration	– (The authors did not introduced them)
Yadav et al. [32]	WEKA	26 cities situated in different climatic zones of India	Global solar radiation	Latitude, longitude, altitude, sunshine duration as well as maximum, minimum and average air temperatures	Maximum, minimum and average air temperatures as well as altitude and sunshine duration
Mohammadi et al. [33]	SVM-WT	Bandar Abass, Iran	Global solar radiation	Three different input combinations of relative sunshine duration, difference between maximum and minimum air temperatures, relative humidity, water vapor pressure, average air temperature and extraterrestrial solar radiation	Combination of relative sunshine duration, difference between maximum and minimum air temperatures, relative humidity, water vapor pressure, average air temperature and extraterrestrial solar radiation
Yadav and Chandel [34]	J48 algorithm using WEKA	26 locations in Himachal Pradesh state, India	Global solar radiation	Air temperature, altitude and sunshine duration	Air temperature, altitude, sunshine duration, latitude, longitude, clearness index and extraterrestrial radiation
Lopez et al. [35]	ARD	A desert location in USA	Direct solar radiation	Air temperature, relative humidity, atmospheric pressure, dew point temperature, precipitable water, solar zenith angle, clearness index, relative air mass and wind speed	clearness index and relative air mass
Mohammadi et al. [36]	ANFIS	Kerman, Iran	Diffuse solar radiation	Global solar radiation, sunshine duration, maximum possible sunshine duration, maximum, minimum and average air temperatures, relative humidity, water vapor pressure, solar declination angle, extraterrestrial solar radiation	For the cases with one input, sunshine duration was determined as the most important variable. For the cases with 2 and 3 inputs, combination of global solar radiation and extraterrestrial solar radiation as well as combination of global solar radiation and extraterrestrial solar radiation and sunshine duration

region. Basically, there can be drawbacks in the inclusion of many input variables. Some of the drawbacks would include the difficulty in explaining the model, inaccuracies caused by irrelevant parameters, complexity in the developed model due to high number of required inputs and time consuming task for collecting more data. These factors may consequently deteriorate the generalization capacity of the model.

This process of selection is usually named variable selection. Essentially, with neural network (NN) as the foundation, the complex system's architecture in function of approximation and regression is modeled. This is achieved in the process of variable selection. The purpose of this procedure is to find a subset of the total set of parameters that have been recorded to show favorable capability of prediction [37–41]. Among many NN systems, the adaptive neuro-fuzzy inference system (ANFIS) is one of the most utilized and powerful techniques. Thus, in this study the ANFIS is employed for the first time to select the most influential variables on global solar radiation prediction [42,43].

To determine how the nine above-mentioned parameters influence the global solar radiation, a parameter search by employing the ANFIS was conducted. ANFIS [44] is a hybrid intelligent system which increases the capability of learning and adapting automatically, has been widely utilized by researchers in a variety of engineering systems for different purposes such as modeling [45–47], prediction [49–51] and control [52–54,56]. This neuro-adaptive learning methodology allows the fuzzy modeling process to obtain information regarding the data gathered [57,58]. This is the foundational idea underlying all neuro-adaptive learning methodologies. The ANFIS methodology aims to organize the FIS (fuzzy inference system) by analyzing the input/output data pairs [59,60]. It gives fuzzy logic the ability to adjust the MF parameters so that it is optimal in allowing the associated FIS to detect and trace the given input/output data [61,62].

3. Case studies and data sets

Iran is a vast country situated in Middle East with an area of 1,648,195 km^2. Iran enjoys highly diverse climatic conditions with very high temperature difference between the hottest and coldest regions of the country. Despite its considerable climate diversity, Iran is one of the countries which enjoys favorable level of solar radiation and sunshine hours. Central and south central parts of Iran are placed among the sunniest parts of the country and even the world. For this study, three cities of Isfahan, Kerman and Tabass distributed in central and south central parts have been considered as case studies. As all three cities are located in sunny regions, they enjoy remarkable solar energy potential during the entire year. The geographical locations of Isfahan, Kerman and Tabass are presented in Table 2.

The city of Isfahan, as capital of Isfahan province, is located in the semi-dessert region in the central part of Iran. Isfahan enjoys mild and dry weather condition with low level of participation throughout the year, which is only around 123 mm [63]. On the basis of the Köppen classification, the climate of Isfahan is categorized as BWk, which relates to arid desert cold [64]. The city of Kerman, as the center of Kerman province, is the second case study. Kerman is situated in the south central part of Iran with a semi-moderate and dry climate [65]. Based upon the Köppen classification, the Kerman climate condition is also classified as BWk [64]. The city of Tabass, as the third nominated case study, is placed in the central desert of Iran in the South-Khorasan province. Tabass climate is characterized with hot summers and rare snowfall in the winters [66]. Also, according to the Köppen classification its climate is categorized as BWh, which relates to arid desert hot [64].

Table 2
Geographical locations of Isfahan, Kerman and Tabass.

City	Latitude (°N)	Longitude (°E)	Elevation from sea level (m)
Isfahan	32°37′	51°40′	1550
Kerman	30°15′	56°58′	1754
Tabass	33°36′	56°55′	711

Table 3
Information of the utilized data for training and testing phases.

City	Period of used data for training	Period of used data for testing
Isfahan	1988–1991 & 1998–2000	2001–2003
Kerman	1995–2001	2002–2004
Tabass	1998–2002	2003–2004

For this research work, long-term measured data provided by Iranian Meteorological Organization (IMO) have been utilized. The used data sets for all cities include measured daily global solar radiation on a horizontal surface (H), sunshine duration (n), average ambient temperature (T_{avg}), maximum ambient temperature (T_{max}), minimum ambient temperature (T_{min}), relative humidity (R_h) and sea level pressure (P) as well as the calculated values of daily water vapor pressure (V_P), daily maximum possible sunshine duration (N) and extraterrestrial solar radiation on a horizontal surface (H_o). It is worth mentioning that for most Iranian meteorological stations including the considered case studies, the water vapor pressure is calculated based upon the measurements of other parameters. In fact, in the data sets used for this study, the amount of water vapor pressure were computed and offered by the meteorological experts and staff working in IMO. Moreover, the values of N and H_o were computed using the equations presented in the Appendix 1.

Due to some missing and also unreliable values in the used daily global solar radiation data, prior to performing any attempt for data analysis, the quality of global solar radiation data was enhanced. For this aim, first, the daily clearness index (K_T) was computed and the values which were out of range of $0.015 < K_T < 1$ were eliminated [67,68]. Second, if in a month there were a few missing or unreliable global solar radiation values they were substituted by proper values obtained using interpolation [67,68].

The available data for this study were divided into two parts of training and testing data sets. Table 3 presents the time periods of data used for training and testing for Isfahan, Kerman and Tabass. It should be mentioned that these are the only available reliable horizontal global solar radiation data for selected case studies. Furthermore, for city of Isfahan for the period of 1992–1997 which were not included in this study, the horizontal global solar radiation data contained either so many missing or unreliable values. Thus, this period of data was eliminated from the analysis.

4. Simulation using ANFIS technique

In this section, the process of conducted simulation using ANFIS technique to select the more relevant variables for global solar radiation prediction is explained. On this account, first, the input and output variables utilized for modeling are offered in the following sub-section.

4.1. Input and output variables

As stated earlier, identifying the most significant parameters which are potentially influential in predicting the daily horizontal global solar radiation (H) is the main objective of this research work. Nine widely available input elements including n (input 1), N (input 2), T_{min} (input 3), T_{max} (input 4), T_{avg} (input 5), R_h (input 6), V_P (input 7), P (input 8) and H_o (input 9) were considered to assess their influence on prediction of H (output). Table 4 presents some descriptive statistics including mean values, standard deviation, minimum and maximum values as well as the range of the 9 input parameters for all selected case studies. Also, similar descriptive statistics for H as the output parameter are offered in Table 5.

4.2. Variable selection using ANFIS

Generating predetermined input-output subsets requires the construction of a set of fuzzy 'IF THEN' rules with the suitable MFs (membership function). The ANFIS can serve as the foundation for such a construction. The input-output data are converted membership functions. In accordance to the collection of input-output data, the ANFIS takes the initial FIS and adjusts it through a back propagation algorithm. The FIS consists of three components: (1) a rule base, (2) a database and (3) a reasoning mechanism. The rule base includes a choice of fuzzy rules. The database assigns the MFs which are employed in the fuzzy rules. Finally, the last component is the reasoning mechanism and it infers from the rules and input data to come to a feasible outcome. These intelligent systems are a combination of knowledge, methods and techniques from a variety of different sources. They are adjusted to perform better in the environments which are changing. These systems have similar-human intelligence within a specific domain. The ANFIS recognizes patterns and assists in the revision of environments. FIS integrates human comprehension, does interfacing, and makes decisions.

FIS in MATLAB is employed in the whole process of the FIS training and evaluation. Fig. 1 illustrates a simple flow chart of the ANFIS selection procedure. Also, Fig. 2 shows all inputs parameters in the ANFIS selection procedure. The ANFIS model should select the parameter or a set of parameters which are the most influential to the output (i.e. daily horizontal global solar radiation). For this aim, three cases with 1, 2 and 3 input parameters are considered and evaluated. As a sample, an ANFIS network for 2 input variables is depicted in Fig. 3 and then its description is presented in the following.

The fuzzy IF-THEN rules of Takagi and Sugeno's class and two inputs for the first-order Sugeno are employed in this study:

$$\text{if } x \text{ is } A \text{ and } y \text{ is } C \text{ then } f_1 = p_1 x + q_1 y + r_1 \tag{1}$$

The 1st layer is made up of input parameters MFs, and it provides the input values to the following layer. Each node is considered as an adaptive node having a node function $O = \mu_{AB}(x)$ and $O = \mu_{CD}(x)$ where $\mu_{AB}(x)$ and $\mu_{CD}(x)$ are membership functions. Bell-shaped membership functions having the maximum value (1.0) and the minimum value (0.0) are selected, such as:

$$\mu(x) = bell(x; a_i, b_i, c_i) = \frac{1}{1 + \left[\left(\frac{x - c_i}{a_i}\right)^2\right]^{b_i}} \tag{2}$$

where $\{a_i, b_i, c_i\}$ is the set of parameters set. The parameters of this layer are designated as premise parameters. Here, x and y are the inputs to nodes.

The membership layer is the second layer. It looks for the weights of every membership function. This layer gets the receiving signals from the preceding layer and then it acts as membership function to the representation of the fuzzy sets of each input variable, respectively. Second layer nodes are non-adaptive. The layer acts as a multiplier for the receiving signals and

Table 5
Descriptive statistics for daily horizontal global solar radiation as output parameter.

Parameter	City	H (MJ/m^2)
Min	Isfahan	1.2
Max		34.7
Mean		19.5
St. dev.		7.1
Range		33.5
Min	Kerman	0.9
Max		33.5
Mean		20.6
St. dev.		7.0
Range		32.6
Min	Tabass	1.3
Max		37.6
Mean		20.2
St. dev.		7.2
Range		36.4

Fig. 1. Simple flow chart of the ANFIS selection procedure.

Table 4
Descriptive statistics for the utilized daily input parameters.

Parameter	City	n (hr)	N (hr)	T_{min} (°C)	T_{max} (°C)	T_{avg} (°C)	R_h (%)	V_P (mb)	P (mb)	H_o (MJ/m^2)
Min	Isfahan	0.8	9.9	−14.4	−1.0	−5.2	7.4	1.1	986.5	18.2
Max		13.9	14.1	28.2	43.0	34.4	98.8	18.0	1036.2	41.4
Mean		9.2	12.0	9.7	24.3	17.0	35.2	6.3	1009.9	30.8
St. dev.		3.0	1.5	8.8	10.2	9.3	17.1	2.6	9.3	8.3
Range		13.1	4.2	42.6	44.0	39.6	91.4	16.9	49.7	23.2
Min	Kerman	0.5	10.1	−17.2	−4.0	−10.1	7.0	1.0	987.2	19.6
Max		13.4	13.9	26.6	42.0	32.2	98.4	16.1	1035.6	41.2
Mean		9.1	12.0	7.1	25.1	16.1	32.0	5.7	1009.8	31.4
St. dev.		3.2	1.4	8.1	9.0	8.2	17.2	2.2	9.3	7.7
Range		12.9	3.8	43.8	46.0	42.3	91.4	15.1	48.4	21.6
Min	Tabass	0.5	9.8	−5.2	2.2	0.7	9.4	1.8	989.7	17.6
Max		13.4	14.2	35.4	48.6	40.5	93.9	20.6	1037.1	41.5
Mean		9.1	12.0	15.7	28.4	22.0	32.2	7.3	1011.2	30.5
St. dev.		3.2	1.5	9.6	10.9	10.1	17.7	2.6	9.4	8.5
Range		12.9	4.4	40.6	46.4	39.8	84.5	18.7	47.4	23.9

sends out the outcome in $w_i = \mu_{AB}(x) * \mu_{CD}(y)$ form. Every output node exhibits the firing strength of a rule.

The third layer is known as the rule layer. All neurons in this layer act as the pre-condition matching the fuzzy rules i.e. each rule's activation level is calculated whereby the number of fuzzy rules is equal to the quantity of layers. Every node calculates the normalized weights. The nodes in the 3rd layer are also considered non-adaptive. Each node computes the value of the rule's firing strength over the sum of all rules' firing strengths in the form of $w_i^* = \frac{w_i}{w_1 + w_2}$, $i = 1, 2$. The outcomes are referred to as the normalized firing strengths.

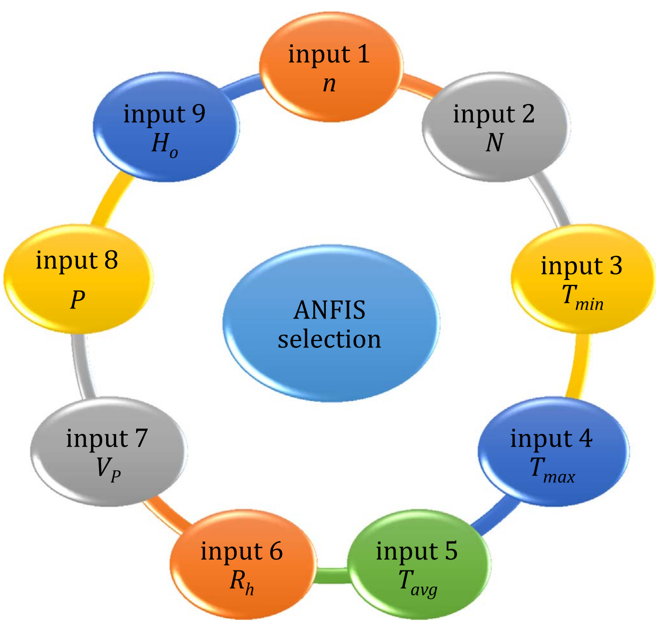

Fig. 2. Input parameters for ANFIS selection procedure.

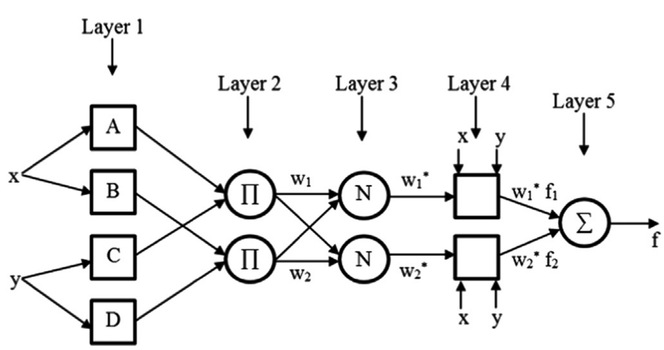

Fig. 3. ANFIS structure with two inputs.

The 4th layer provides the output values as a result of the inference of rules. This layer is known as the defuzzification layer. Every 4th layer node is an adaptive node having the node function $O_i^4 = w_i^* xf = w_i^*(p_i x + q_i y + r_i)$. In this layer, the $\{p_i, q_i, r\}$ is the variable set. The variable set is designated as the consequent parameters.

The 5th layer (final layer) is known as the output layer. It adds up all the receiving inputs from the preceding layer. Thereafter, it converts the fuzzy classification outcomes into a binary (crisp). The single node of the 5th layer is considered non-adaptive. This node calculates the total output as the whole sum of all receiving signals,

$$O_i^5 = \sum_i w_i^* f = \frac{\sum_i w_i f}{\sum_i w_i} \tag{3}$$

In the process of identification of variables in the ANFIS architectures, the hybrid learning algorithms are applied. The functional signals progress until the 4th layer whereby the hybrid learning algorithm passes. Further, the consequent variables are found by the least squares estimation. In the backward pass, the error rates circulate backwards and the premise variables are synchronized through the gradient decline order.

5. Results and discussion

In this study, using the considered data sets and ANFIS technique, an extensive exploration was performed to evaluate the influence of nine important input variables of n, N, T_{min}, T_{max}, T_{avg}, R_h, V_P, P and H_o on prediction of daily horizontal global solar radiation (H) and then identify the most relevant sets of input parameters. 50% of the collected data sets were used for training and the subsequent 50% served for testing. Several ANFIS models were built which include functions for various possible combinations of the used inputs. Then they were trained respectively for single epoch and then the achieved performance on the basis of each combination was reported. Three statistical parameters of mean absolute bias error (MABE), root mean square error (RMSE) and correlation coefficient (R) were used as reliable indicators to show the accuracy level of global solar radiation prediction using each combination of inputs and subsequently determine the rank of each input parameter from the most relevant to least relevant. The brief descriptions of these statistical parameters are given in Appendix 2. The input or set of inputs which provides the lowest errors and highest R is considered as the most relevant variable for daily global solar radiation prediction. The selection of most significant parameters was carried out for three cases. For the first case, only one input was considered and the rank of inputs was identified. Furthermore, for the second and third cases different possible combinations of parameters with respectively two and three input elements were considered. Afterwards the best and worst sets of inputs were determined. The examinations were

Table 6
Achieved statistical parameters for both training (tra) and testing (tes) periods and rank of input parameters for case of using 1 input for Isfahan.

ANFIS model	Input	MABE (tra)	MABE (tes)	RMSE (tra)	RMSE (tes)	R (tra)	R (tes)	Rank
Model (1)	n (input 1)	2.2397	2.1187	3.6175	3.5476	0.9875	0.9891	1
Model (2)	N (input 2)	2.4556	2.4023	3.8374	3.6921	0.9812	0.9819	3
Model (3)	T_{min} (input 3)	3.2223	3.1247	4.8469	4.6458	0.9324	0.9345	8
Model (4)	T_{max} (input 4)	2.9112	2.8842	4.3817	4.2659	0.9478	0.9489	4
Model (5)	T_{avg} (input 5)	3.1153	3.0444	4.5254	4.3642	0.9406	0.9427	5
Model (6)	R_h (input 6)	3.1723	3.1191	4.5779	4.5961	0.9388	0.9376	6
Model (7)	V_P (input 7)	4.5643	4.5128	6.7459	6.6297	0.8723	0.8745	9
Model (8)	P (input 8)	3.3221	3.2119	4.7440	4.6089	0.9345	0.9352	7
Model (9)	H_o (input 9)	2.3176	2.2937	3.6256	3.5624	0.9843	0.9853	2

Table 7
Achieved statistical parameters for both training (tra) and testing (tes) periods and rank of input parameters for case of using 1 input for Kerman.

ANFIS model	Input	MABE (tra)	MABE (tes)	RMSE (tra)	RMSE (tes)	R (tra)	R (tes)	Rank
Model (1)	n (input 1)	2.6613	2.6121	4.0102	3.8910	0.9762	0.9783	1
Model (2)	N (input 2)	2.8143	2.8186	4.4152	4.4238	0.9461	0.9469	3
Model (3)	T_{min} (input 3)	3.2734	3.2698	5.1891	5.1788	0.9211	0.9249	7
Model (4)	T_{max} (input 4)	2.8456	2.8512	4.4455	4.4622	0.9432	0.9419	4
Model (5)	T_{avg} (input 5)	2.9324	2.9436	4.5798	4.5971	0.9423	0.9398	5
Model (6)	R_h (input 6)	3.2213	3.2422	5.1359	5.1888	0.9277	0.9238	8
Model (7)	V_P (input 7)	4.2278	4.2266	6.7008	6.6995	0.8766	0.8787	9
Model (8)	P (input 8)	3.0980	3.1211	4.7292	4.8096	0.9355	0.9323	6
Model (9)	H_o (input 9)	2.7235	2.7023	4.0400	3.9198	0.9733	0.9741	2

Table 8
Achieved statistical parameters for both training (tra) and testing (tes) periods and rank of input parameters for case of using 1 input for Tabass.

ANFIS model	Input	MABE (tra)	MABE (tes)	RMSE (tra)	RMSE (tes)	R (tra)	R (tes)	Rank
Model (1)	n (input 1)	2.4386	2.4113	3.8171	3.7682	0.9812	0.9831	3
Model (2)	N (input 2)	2.2235	2.2018	3.6007	3.4937	0.9887	0.9892	1
Model (3)	T_{min} (input 3)	2.6923	2.6876	4.1267	4.1041	0.9718	0.9733	4
Model (4)	T_{max} (input 4)	2.7591	2.7529	4.3125	4.2770	0.9564	9.9579	6
Model (5)	T_{avg} (input 5)	2.7163	2.7112	4.1659	4.1314	0.9539	0.9512	5
Model (6)	R_h (input 6)	3.3002	3.2989	5.1946	5.1806	0.9266	9.9283	8
Model (7)	V_P (input 7)	4.1887	4.1809	6.3906	6.2599	0.8875	0.8913	9
Model (8)	P (input 8)	2.8439	2.8112	4.5237	4.4349	0.9474	0.9433	7
Model (9)	H_o (input 9)	2.2413	2.2399	3.6318	3.5247	0.9865	0.9882	2

performed separately for each considered case study to offer further reliability in the evaluations since each case study enjoys different climate conditions and solar radiation characteristics. The attained results are presented and discussed in the following sub-sections.

5.1. Case 1: parameter selection for 1 input

For the first case, only one parameter is considered as input. For this purpose, 9 ANFIS models were developed to evaluate the importance of each input variable and identify the most relevant input. The significance of inputs is ranked based upon the considered statistical parameters of MABE (MJ/m^2), RMSE (MJ/m^2) and R (dimensionless). Tables 6–8 present the results of ANFIS regression errors for both training and testing phases as well as the achieved rank of input parameters, respectively for Isfahan, Kerman and Tabass. The obtained results reveal that for Isfahan and Kerman the variable of sunshine hour (n or input 1) and for Tabass the variable of maximum possible sunshine hour (N or input 2) are identified as the most relevant parameters for prediction of global solar radiation. It is also found that for all cities n (input 1), N (input 2) and H_o (input 9) are the most influential variables.

Basically, n and N have been extensively endorsed as highly effective variables for prediction of global solar radiation (H) [69]. Based on a particular definition of the World Meteorological Organization (WMO), n is the sum of periods for which the beam (direct) solar irradiance exceeds 120 W/m^2 [70]. It typically defines the sky conditions in every location as clear sky, partly cloudy sky and cloudy sky conditions. The N, or the expected day length, is the time interval between sunrise and sunset. H_o, as another important element for prediction of global solar radiation, is a constant value for each specific day in any geographical location, irrespective of change of year. Nevertheless, solar attenuation happens as radiation passes through the atmosphere due to different atmospheric phenomenon. The ratio between the H and H_o defines the concept of clearness index (K_T). The Angström–Prescott model is a widely used regression based global solar radiation model [7,8]. This model has been successfully utilized to predict horizontal global solar radiation in many locations around the globe. It has been developed based on the correlation of H with n, N and H_o. The successful application and high flexibility of this model among several regression based global solar radiation models highlight the significance of n, N and H_o on prediction of H, as the achieved results of this study further prove this point.

Also, V_P (input 7) is recognized as the least relevant parameter for all cities. The difference between the utilized statistical parameters for the considered variables indicates the importance of proper selection of input parameters. This difference is indeed significant between the most and least relevant input parameters. The fact that both the training and testing statistical parameters are comparable is an indirect indication that suggests that there is no over fitting. This means that the selection of more than one input parameter in the construction of the ANFIS model can be explored. Thus, in the Sections 5.2 and 5.3 the most significant combinations of parameters using 2 and 3 inputs are determined and discussed, respectively. It should be highlighted that the least relevant input (V_P or input 7) is not omitted from the possible sets of parameters to provide and assess the importance of more combinations of inputs.

5.2. Case 2: parameter selection for combination of 2 inputs

To recognize the best integration of 2 parameters, 36 possible sets of inputs (36 ANFIS models) were considered and analyzed. Table 9 provides comparisons between the most and least relevant combinations of 2 input variables and the achieved values of statistical parameters for both training and testing phases for Isfahan, Kerman and Tabass. Obviously, once the lowest errors and highest R are attained for a set of inputs, it is determined as the most relevant combination of 2 inputs and vice versa. It is evident that the optimal combination of inputs is different among cities due to different solar radiation characteristics and weather conditions. For Isfahan the combination of n and H_o (inputs 1 and 9) and for Kerman and Tabass the combination of n and N (inputs 1 and 2)

Table 9
The best and worst combinations of 2 inputs and related statistical parameters for both training (tra) and testing (tes) periods for Isfahan, Kerman and Tabass.

City	Best/worst model	Combination of inputs	MABE (tra)	MABE (tes)	RMSE (tra)	RMSE (tes)	R (tra)	R (tes)
Isfahan	Best	n and H_o (inputs 1 and 9)	1.6533	1.6488	2.4899	2.3596	0.9922	0.9931
	Worst	V_P and P (inputs 7 and 8)	2.8566	2.8502	4.6139	4.5120	0.9455	0.9463
Kerman	Best	n and N (inputs 1 and 2)	1.8835	1.8854	2.7337	2.7485	0.9896	0.9892
	Worst	T_{min} and P (inputs 3 and 8)	2.9932	3.0051	4.7267	4.7742	0.9421	0.9403
Tabass	Best	n and N (inputs 1 and 2)	2.1139	2.1103	3.4263	3.3200	0.9877	0.9886
	Worst	V_P and P (inputs 7 and 8)	2.7869	2.7478	4.5133	4.4288	0.9438	0.9459

Table 10
The best and worst combinations of 3 inputs and related statistical parameters for both training (tra) and testing (tes) periods for Isfahan, Kerman and Tabass.

City	Best/worst model	Combination of inputs	MABE (tra)	MABE (tes)	RMSE (tra)	RMSE (tes)	R (tra)	R (tes)
Isfahan	Best	n, N and T_{max} (inputs 1, 2 and 4)	1.4839	1.4718	2.4232	2.3537	0.9929	0.9932
	Worst	T_{min}, T_{max} and T_{avg} (inputs 3, 4 and 5)	2.6881	2.6711	4.1938	4.1377	0.9530	0.9588
Kerman	Best	n, N and T_{max} (inputs 1, 2 and 4)	1.7553	1.7690	2.6527	2.7345	0.9899	0.9878
	Worst	T_{avg}, V_P and P (inputs 3, 7 and 8)	2.7223	2.7672	4.3804	4.5220	0.9476	0.9455
Tabass	Best	n, N and H_o (inputs 1, 2 and 9)	2.1227	2.1176	3.4033	3.3122	0.9871	0.9880
	Worst	T_{max}, R_h and V_P (inputs 4, 6 and 7)	2.6569	2.6703	4.0613	4.1526	0.9551	0.9537

are determined as the most significant sets of 2 inputs. Clearly, it is found that in the optimal sets of inputs for each city there are two out of three parameters of n, N and H_o determined as the most influential elements. Owing to the wide accessibility of the recorded sunshine duration in most meteorological stations, using these combinations of inputs would be highly applicable to predict global solar radiation.

Although identifying the best combinations of input parameters is of particular significance, having knowledge on worst combination inputs may also be interesting. It is found that combinations of T_{min} and P (inputs 3 and 8) as well as V_P and P (inputs 7 and 8) as inputs are among the worst choice to predict the global solar radiation. The comparison between the highest and lowest rank (i.e. the best and least combinations of inputs) is profitable to demonstrate that how much difference in predictions accuracy exists between most and least relevant sets of 2 inputs. For Isfahan and Kerman these differences are higher than Tabass.

5.3. Case 3: parameter selection for combination of 3 inputs

Final test was performed to select the most important sets combined of 3 input parameters. To achieve this, 84 different combinations of 3 inputs (84 ANFIS models) were considered to determine and introduce the most significant one for prediction of global solar radiation. In Table 10, the most and least significant combinations of 3 input variables and also the attained statistical parameters both training and testing phases are compared for each case study. It is observed that the most influential set of 3 inputs are identical for Isfahan and Kerman while it is different for Tabass. In fact, for Isfahan and Kerman the combination of n, N and T_{max} (inputs 1, 2 and 4) and for Tabass the combination of n, N and H_o (inputs 1, 2 and 9) are recognized as the optimal combinations of 3 inputs. The attained results for the least significant combinations of 3 parameters indicate that the worst combinations of inputs are different for cities. It is noticed that the worst combination of 3 inputs for Isfahan is the combination of T_{min}, T_{max} and T_{avg} (inputs 3, 4 and 5) while for Kerman and Tabass are T_{avg}, V_P and P (inputs 3, 7 and 8) and also T_{max}, R_h and V_P (inputs 4, 6 and 7), respectively. The differences between the used statistical parameters for the best and the worst sets of inputs highlight the importance of appropriate selection of parameters to predict the daily global solar radiation.

5.4. Comparing the cases and introducing the best combination of inputs

Although in the sub-Sections 3.1–3.3 the most significant combinations of 1, 2 and 3 inputs were identified for global solar radiation prediction, it would be important to introduce the best possibility among all. This evaluation is offered here only based on RMSE. Fig. 4(a)–(c) show the amount of RMSE (MJ/m^2) during training and testing phases for the most relevant set of parameters with 1, 2 and 3 inputs, respectively for Isfahan, Kerman and Tabass. It is found that by increasing the number of most relevant input from 1 to 2 and then from 2 to 3, the amount of errors decrease; thus, the prediction accuracy enhance. This improvement is notable for Isfahan and Kerman and it is acceptable for Tabass when the number of inputs increase from 1 to 2. Nevertheless, this enhancement declines significantly for all cities when the number of inputs increases from 2 to 3. Thus, increasing the number of inputs to higher than 3 may not advisable due to minor improvements achieved and further complexity in the required inputs to predict global solar radiation. Generally, as a model with more simplicity in terms of required inputs is always preferable, use of more than two inputs in the construction of the ANFIS model may not be desirable and appropriate. Therefore, considering the most relevant combinations of 2 inputs would be the more suitable possibility in terms of optimum number of inputs to provide a balance between the simplicity and high precision.

Fig. 5(a)–(c) show the ANFIS decision surfaces for the optimal combinations of the two parameters for each city. These surfaces are useful to illustrate that how the output (i.e. daily global solar radiation) in each city varies with the 2 input parameters. According to the Fig. 5(a) and (b), it is interesting to note that the parameters of n and H_o have the similar influence on the global solar radiation as the parameters of n and N. While comparison between Fig. 5(b) and (c) reveals that although n and N are the optimum input parameters for Kerman and Tabass, the variation of H with the parameters of n and N are totally different. This could be due to difference between the climate conditions and solar radiation characteristics of the cities.

6. Conclusions

Accessibility to precise global solar radiation data is of indispensable significance in various applications of solar energy.

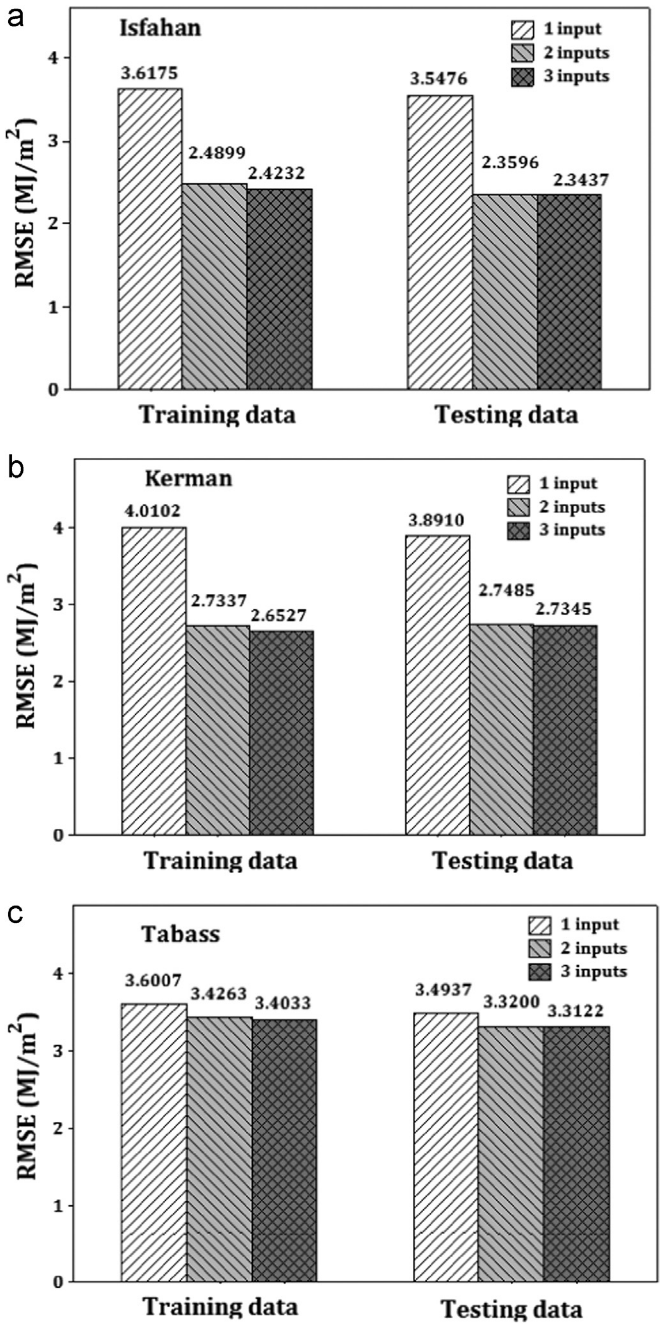

Fig. 4. RMSE (MJ/m^2) for the most significant combinations of parameters with 1, 2 and 3 inputs for (a) Isfahan, (b) Kerman and (c) Tabass.

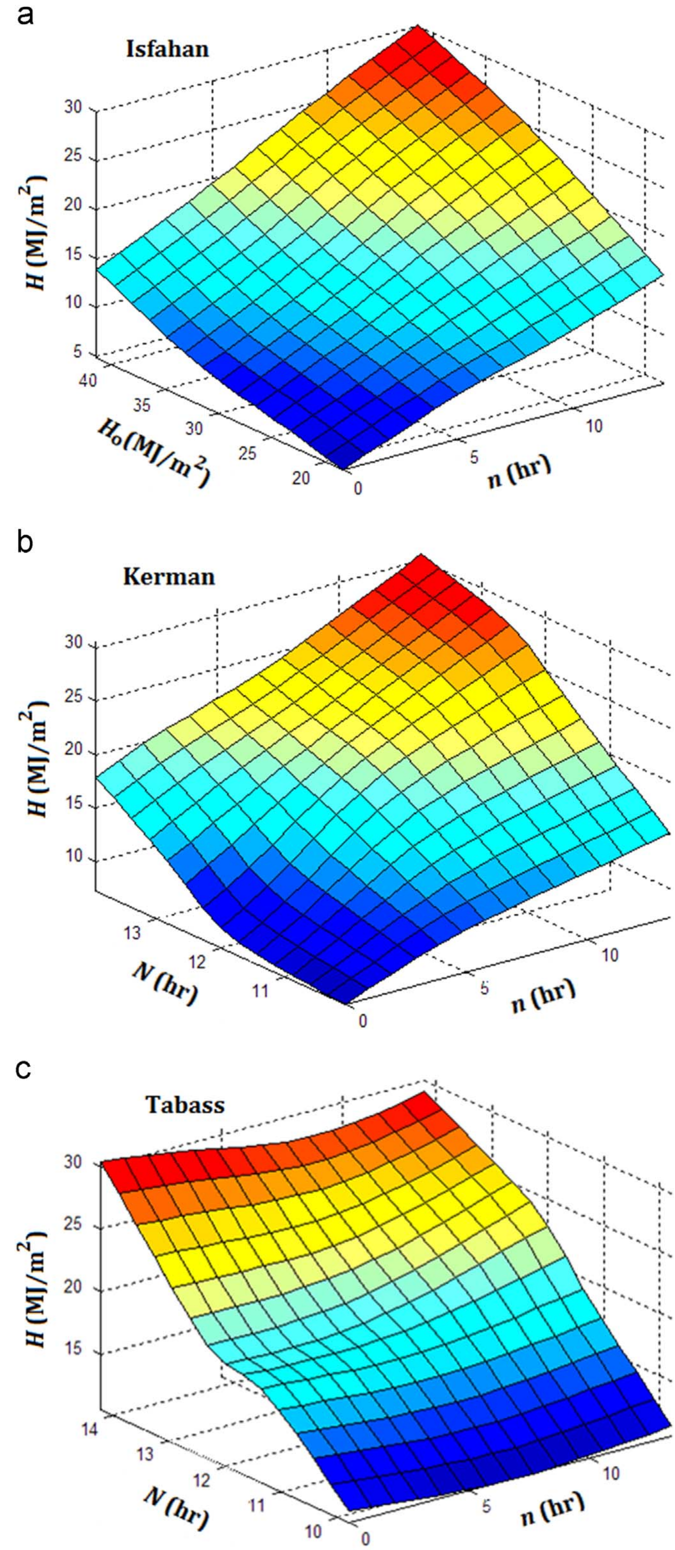

Fig. 5. ANFIS decision surfaces for the optimal combinations of the two the most influential parameters for (a) Isfahan, (b) Kerman and (c) Tabass.

Therefore, determining the most important parameters to predict global solar radiation by means of efficient models and techniques is truly essential. In the study, using ANFIS methodology, a systematic approach was applied to select the most significant parameters for prediction of daily global solar radiation (H). The ANFIS network was used to perform a variable search for determining that how nine important parameters of n, N, T_{min}, T_{max}, T_{avg}, R_h, V_P, P and H_o influence H. Using the given data sets, an extensive examination was performed to select the set of the optimal combinations of parameters with 1, 2 and 3 inputs that have the most influence on prediction of daily global solar radiation. To recognize the best integration of parameters with 1, 2 and 3 inputs, respectively 9, 36 and 84 possible combinations of inputs (9, 36 and 84 ANFIS models) were considered. MABE, RMSE and R were used as reliable indicators to show the accuracy level of prediction using each combination and subsequently determine the rank of each input variable from the most relevant to least relevant set.

It was found that proper selections of input parameters have significant impacts on accurate prediction of daily global solar radiation. Nevertheless, the results indicated that it is not possible to introduce an optimal combination of inputs for all cities because the climate conditions and solar radiation characteristics of the cities are different. For Isfahan and Kerman n and for Tabass N were the most relevant parameters while V_P was the least relevant parameter for all cities. In terms of combination of 2 inputs, for Isfahan n and H_o and for Kerman and Tabass n and N were the most significant sets. Moreover, regarding the optimal set of 3 inputs, for Isfahan and Kerman n, N and T_{max} and for Tabass n, N and H_o were the most optimum combinations. The study results revealed that considering the most relevant combinations of 2 inputs would be regarded as the more suitable possibility to provide a balance between lower simplicity and higher precision in prediction of daily global solar radiation. In fact, these more appropriate combinations of 2 input parameters offer higher accuracy and insights in the relevance of global solar radiation prediction in the selected case studies.

Acknowledgment

This research is supported by University of Malaya under UMRG grant (Project No: RP006A-14HNE).

Appendix 1

The extraterrestrial solar radiation on a horizontal surface (H_o) is expressed as [69]:

$$H_o = \frac{24 \times 3600}{\pi} G_{sc}\left(1 + 0.033\cos\frac{360\,n_{day}}{365}\right) \times \left(\cos\varphi\cos\delta\sin\omega_s + \frac{\pi\omega_s}{180}\sin\varphi\sin\delta\right)$$

where G_{sc} is the solar constant, assumed equal to 1367 W/m^2 and n_{day} is the average day of each month [70]. δ and ω_s are the daily solar declination and sunset hour angles, respectively, as [69]:

$$\delta = 23.45\sin\left(\frac{(n_{day}+284)360}{365}\right)$$

$$\omega_s = \cos^{-1}(-\tan\varphi\tan\delta)$$

The maximum possible sunshine duration (N) is [69]:

$$N = \frac{2}{15}\cos^{-1}(-\tan\varphi\tan\delta)$$

Appendix 2

The mean absolute bias error (MABE), which indicate the average quantity of total absolute bias errors between predicted and measured values, is given by [71]:

$$MABE = \frac{1}{x}\sum_{i=1}^{x}\left|H_{i,c} - H_{i,m}\right|$$

where x is the total number of data.

The root mean square error (RMSE), which determines the model's accuracy by comparing the deviation between the predicted and measured values, is defined as [71]:

$$RMSE = \sqrt{\frac{1}{x}\sum_{i=1}^{x}(H_{i,c} - H_{i,m})^2}$$

The correlation coefficient, R, utilized for testing the linear relationship of the predicted data with measured data is defined by [71]:

$$R = \frac{\sum_{i=1}^{x}(H_{i,c} - H_{c,a\,vg})\cdot(H_{i,m} - H_{m,avg})}{\sqrt{\left[\sum_{i=1}^{x}(H_{i,c} - H_{c,av\,g})^2\right]\left[\sum_{i=1}^{x}(H_{i,m} - H_{m,av\,g})^2\right]}}$$

where $H_{c,avg}$ and $H_{m,avg}$ are the averaged predicted and measured values, respectively.

References

[1] Wu Yujie, Wang Jianzhou. A novel hybrid model based on artificial neural networks for solar radiation prediction. Renew Energy 2016;89:268–84.
[2] Boland J, Huang J, Ridley B. Decomposing global solar radiation into its direct and diffuse components. Renew Sustain Energy Rev 2013;28:749–56.
[3] Despotovic Milan, Nedic Vladimir, Despotovic Danijela, Cvetanovic Slobodan. Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation. Renew Sustain Energy Rev 2016;56:246–60.
[4] Demirhan Haydar, Atilgan Yasemin Kayhan. New horizontal global solar radiation estimation models for Turkey based on robust coplot supported genetic programming technique. Energy Convers Manag 2015;106:1013–23.
[5] Olatomiwa Lanre, Mekhilef Saad, Shamshirband Shahaboddin, Petković Dalibor. Adaptive neuro-fuzzy approach for solar radiation prediction in Nigeria. Renew Sustain Energy Rev 2015;51:1784–91.
[6] Gairaa Kacem, Khellaf Abdallah, Messlem Youcef, Chellali Farouk. Estimation of the daily global solar radiation based on Box–Jenkins and ANN models: a combined approach. Renew Sustain Energy Rev 2016;57:238–49.
[7] Angström A. Solar and terrestrial radiation. Q J R Meteorol Soc 1924;50:121–5.
[8] Prescott JA. Evaporation from a water surface in relation to solar radiation. Trans R Soc Sci Aust 1940;64:114–25.
[9] Bahel V, Bakhsh H, Srinivasan R. A correlation for estimation of global solar radiation. Energy 1987;12:131–5.
[10] Abdalla YAG. New correlation of global solar radiation with meteorological parameters for Bahrain. Int J Sol Energy 1994;16:111–20.
[11] Garg HP, Garg ST. Prediction of global solar radiation from bright sunshine hours and other meteorological parameters. In: Solar-India, Proceedings of the national solar energy convention. Allied Publishers: New Delhi; 1982. p. 1004–07.
[12] Ojosu JO, Komolafe LK. Models for estimating solar radiation availability in South Western Nigeria. Niger J Sol Energy 1987;6:69–77.
[13] Ododo JC, Sulaiman AT, Aidan J, Yguda MM, Ogbu FA. The importance of maximum air temperature in the parameterization of solar radiation in Nigeria. Renew Energy 1995;6:751–63.
[14] Elagib AA, Mansell MG. New approaches for estimating global solar radiation across Sudan. Energy Convers Manag 2000;41:419–34.
[15] Robaa SM. Validation of the existing models for estimating global solar radiation over Egypt. Energy Convers Manag 2009;50:184–93.
[16] Teke A, Yıldırım HB. Estimating the monthly global solar radiation for Eastern Mediterranean Region. Energy Convers Manag 2014;87:628–35.
[17] Al-Mostafa ZA, Maghrabi AH, Al-Shehria SM. Sunshine-based global radiation models: a review and case study. Energy Convers Manag 2014;84:209–16.
[18] Demirhan H. The problem of multicollinearity in horizontal solar radiation estimation models and a new model for Turkey. Energy Convers Manag 2014;84:334–45.
[19] Kaplanis S, Kumar Jatin, Kaplani E. On a universal model for the prediction of the daily global solar radiation. Renew Energy 2016;91:178–88.
[20] Quej Victor H, Almorox Javier, Ibrakhimov Mirzakhayot, Saito Laurel. Empirical models for estimating daily global solar radiation in Yucatán Peninsula, Mexico. Energy Convers Manag 2016;110:448–56.
[21] Tymvios FS, Jacovides CP, Michaelides SC, Scouteli C. Comparative study of Angströms and artificial neural networks' methodologies in estimating global solar radiation. Sol Energy 2005;78:752–62.
[22] Bosch JL, López G, Batlles FJ. Daily solar irradiation estimation over a mountainous area using artificial neural networks. Renew Energy 2008;33:1622–8.
[23] Benghanem M, Mellit A, Alamri SN. ANN-based modelling and estimation of daily global solar radiation data: a case study. Energy Convers Manag 2009;50:1644–55.
[24] Behrang MA, Assareh E, Ghanbarzadeh A, Noghrehabadi AR. The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data. Sol Energy 2010;84:1468–80.
[25] Koca A, Oztop HF, Varol Y, Koca GO. Estimation of solar radiation using artificial neural networks with different input parameters for Mediterranean region of Anatolia in Turkey. Expert Syst Appl 2011;38:8756–62.
[26] Yacef R, Benghanem M, Mellit A. Prediction of daily global solar irradiation data using Bayesian neural network: a comparative study. Renew Energy 2012;48:146–54.
[27] Ozgoren M, Bilgili M, Sahin B. Estimation of global solar radiation using ANN over Turkey. Expert Syst Appl 2012;39:5043–51.
[28] Mostafavi ES, Saeidi Ramiyani S, Sarvar R, Izadi Moud H, Mousavi SM. A hybrid computational approach to estimate solar global radiation: an empirical evidence from Iran. Energy 2013;49:204–10.

[29] Bhardwaj S, Sharma V, Srivastava S, Sastry OS, B. Bandyopadhyay B, Chandel SS, et al. Estimation of solar radiation using a combination of hidden Markov model and generalized Fuzzy model. Sol Energy 2013;93:43–54.
[30] Ramedani Zeynab, Omid Mahmoud, Keyhani Alireza. Modeling solar energy potential in a Tehran Province using artificial neural networks. Int J Green Energy 2013;10(4):427–41.
[31] Will A, Bustos J, Bocco M, Gotaya J, Lamelas C. On the use of niching genetic algorithms for variable selection in solar radiation estimation. Renew Energy 2013;50:168–76.
[32] Yadav AK, Malik Hasmat, Chandel SS. Selection of most relevant input parameters using WEKA for artificial neural network based solar radiation prediction models. Renew Sustain Energy Rev 2014;31:509–19.
[33] Mohammadi Kasra, Shamshirband Shahaboddin, Tong Chong Wen, Arif Muhammad, Petkovic Dalibor, Sudheer Ch. A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation. Energy Convers Manag 2015;92:162–71.
[34] Yadav AK, Chandel SS. Solar energy potential assessment of western Himalayan Indian state of Himachal Pradesh using J48 algorithm of WEKA in ANN based prediction model. Renew Energy 2015;75:675–93.
[35] López G, Batlles FJ, Tovar-Pescador J. Selection of input parameters to model direct solar irradiance by using artificial neural networks. Energy 2005;30:1675–84.
[36] Mohammadi Kasra, Shamshirband Shahaboddin, Petković Dalibor, Khorasanizadeh Hossein. Determining the most important variables for diffuse solar radiation prediction using adaptive neuro-fuzzy methodology; case study: city of Kerman, Iran. Renew Sustain Energy Rev 2016;53:1570–9.
[37] Castellano G, Fanelli AM. Variable selection using neural-network models. Neurocomputing 2000;31:1–13.
[38] Dieterle F, Busche S, Gauglitz G. Growing neural networks for a multivariate calibration and variable selection of time-resolved measurements. Anal Chim Acta 2003;490:71–83.
[39] Cibas T, Soulie FF, Gallinari P, Sarunas Raudys S. Variable selection with neural networks. Neurocomputing 1996;12:223–48.
[40] Anderson FO, Aberg M, Sven P, Jacobsson SP. Algorithmic approaches for studies of variable influence, contribution and selection in neural networks. Chemom Intell Lab Syst 2000;51:61–72.
[41] Donald AS. Using genetic algorithm based variable selection to improve neural network models for real-world systems. In: Proceedings of the 2002 international conference on machine learning & applications; 2002. p. 16–9.
[42] Chan KY, Ling SH, Dillon TS, Nguyen HT. Diagnosis of hypoglycemic episodes using a neural network based rule discovery system. Expert Syst Appl 2011;38(8):9799–808.
[43] Kwong CK, Wong TC, Chan KY. A methodology of generating customer satisfaction models for new product development using a neuro-fuzzy approach. Expert Syst Appl 2009;36(8):11262–70.
[44] Jang J-SR. ANFIS: adaptive-network-based fuzzy inference systems. IEEE Trans Syst Man Cybern 1993;23:665–85.
[45] Ghandoor A, Al, Samhouri M. Electricity consumption in the industrial sector of Jordan: application of multivariate linear regression and adaptive neurofuzzy techniques. Jordan J Mech Ind Eng 2009;3(1):69–76.
[46] Singh R, Kianthola A, Singh TN. Estimation of elastic constant of rocks using an ANFIS approach. Appl Soft Comput 2012;12:40–5.
[47] Vlastimir Nikolić et al. Wind wake influence estimation on energy production of wind farm by adaptive neuro-fuzzy methodology, Energy, 2015; 80(1): 361–72, ISSN 0360-5442.
[49] Hosoz M, Ertunc HM, Bulgurcu H. An adaptive neuro-fuzzy inference system model for predicting the performance of a refrigeration system with a cooling tower. Expert Syst Appl 2011;38:14148–55.
[50] Khajeh A, Modarress H, Rezaee B. Application of adaptive neuro-fuzzy inference system for solubility prediction of carbon dioxide in polymers. Expert Syst Appl 2009;36:5728–32.
[51] Sivakumar R, Balu K. ANFIS based distillation column control. IJCA Spec Issue Evolut Comput Optim Techniq 2010:67–73.
[52] Kurnaz S, Cetin O, Kaynak O. Adaptive neuro-fuzzy inference system based autonomous flight control of unmanned air vehicles. Expert Syst Appl 2010;37:1229–34.
[53] Ravi S, Sudha M, Balakrishnan PA. Design of intelligent self-tuning GA ANFIS temperature controller for plastic extrusion system. Model Simul Eng 2011:1–8.
[54] Areed FG, Haikal AY, Mohammed RH. Adaptive neuro-fuzzy control of an induction motor. Ain Shams Eng J 2010;1:71–8.
[56] Tian L, Collins C. Adaptive neuro-fuzzy control of a flexible manipulator. Mechatronics 2005;15:1305–20.
[57] Aldair AA, Wang WJ. Design an intelligent controller for full vehicle nonlinear active suspension systems. Int J Smart Sens Intell Syst 2011;4(2):224–43.
[58] Dastranj MR, Ebroahimi E, Changizi N, Sameni E, Control DC. Motorspeed with adaptive neuro-fuzzy control (ANFIS). Aust J Basic Appl Sci 2011;5(10):1499–504.
[59] Wahida Banu RSD, Shakila Banu A, Manoj D. Identification and control of nonlinear systems using soft computing techniques. Int J Model Optim 2011;1(1):24–8.
[60] Grigorie TL, Botez RM. Adaptive neuro-fuzzy inference system-based controllers for smart material actuator modelling. J Aerosp Eng 2009:655–68.
[61] Akcayol MA. Application of adaptive neuro-fuzzy controller for SRM. Adv Eng Softw 2004;35:129–37.
[62] Arabic calligraphy: a computational exploration, Hoda Moustapha, School of Architecture, Carnegie Mellon University: Pittsburgh, USA and Ramesh Krishnamurti School of Architecture, Carnegie Mellon University: Pittsburgh, USA; 2001.
[63] 〈http://en.wikipedia.org/wiki/Isfahan〉. [accessed 10.05.15].
[64] Kottek M, Grieser J, Beck C, Rudolf B, Rubel F. World map of the Koppen-Geiger climate classification updated. Meteorol Z 2006;15(3):259–63.
[65] 〈http://en.wikipedia.org/wiki/Kerman〉. [accessed 10.05.15].
[66] 〈http://en.wikipedia.org/wiki/Tabass〉. [accessed 10.05.15].
[67] Shamshirband S, Mohammadi K, Piri J, Petkovic D, Karim A. Hybrid autoregressive neural network model for estimating global solar radiation in Bandar Abbas, Iran. Environ Earth Sci 2016; 75:172, http://dx.doi.org/10.1007/s12665-015-4970-x.
[68] Mohammadi Kasra, Khorasanizadeh Hossein, Shamshirband Shahaboddin, Tong Chong Wen. Influence of introducing various meteorological parameters to the Angström–Prescott model for estimation of global solar radiation. Environ Earth Sci 2016;75:219. http://dx.doi.org/10.1007/s12665-015-4871-z.
[69] Duffie JA, Beckman WA. Solar engineering of thermal processes. 3rd ed. New York: John Wiley & Son; 2006.
[70] Kalogirou SA. Solar energy engineering: processes and systems. 1st ed. London: Elsevier Inc; 2009.
[71] Chang Fi-John, Chang Ya-Ting. Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. Adv Water Resour 2006;29:1–10.